4 ANALYSIS OF FINAL STATUS SURVEY RESULTS: DATA QUALITY ASSESSMENT

Data Quality Assessment (DQA) is the scientific and statistical evaluation of data to determine if the data are of the right type, quality, and quantity to support their intended use (EPA QA/G-9, 1995). There are five steps in the DQA process:

(1) Review the Data Quality Objectives (DQOs) and sampling design.
(2) Conduct a preliminary data review.
(3) Select the statistical test.
(4) Verify the assumptions of the statistical test.
(5) Draw conclusions from the data.

4.1 Review the Data Quality Objectives (DQOs) and Sampling Design

During survey design, acceptable error rates are specified for the statistical tests, and the desired probability that a survey unit will pass the release criterion is charted against the amount of residual radioactivity that actually may be present in order to test the efficacy of a proposed design. During the interpretation of survey results, it is important to determine that the objectives of the design have been met. The first and most straightforward way to check this is to ascertain that the number of usable measurements meet the requirement of the statistical tests as outlined in Section 3.8.1. The sample standard deviation, *s*, should also be compared to the estimate of the measurement variability, σ, that was used to determine the number of samples required. The consequence of there being too few measurements, or of there being higher than expected data variability, is that the Type II error rate β will be larger than planned, and the power of the test to detect departures from the null hypothesis, $1-\beta$, is reduced. In Scenario A this means that a survey unit that meets the release criterion has a higher probability of being incorrectly deemed *not* to meet it. In Scenario B this means that a survey unit that does *not* meet the release criterion has a higher probability of being incorrectly deemed to meet it. After examining the number of usable measurements and their variability, the retrospective power of the test can be determined using the methods of Chapter 10. This is not usually necessary when the null hypothesis is rejected since the Type I error rate, α, is fixed when the statistical test is performed using the actual number of usable measurements.

Since the occurrence of missing or unusable data can impact the Type II error rates, a reasonable allowance for such occurrences should be built into the planning process by adding more measurements to the sample sizes listed in Tables 3.2 and 3.3.

The power of the statistical tests will also be reduced if data variability is greater than that assumed during the survey planning. The number of measurements required to meet the acceptable error specified during the planning process will not be sufficient if σ was underestimated. As mentioned in Section 2.2.6, the overall data variability may be considered to consist of two more or less independent components, the component due to uncertainty in the measurement process, $\sigma_{meas,}$ and the component due to spacial variability in the concentrations being measured, $\sigma_{spatial}$. Spatial variability was discussed in Section 3.5.1. The overall variability is approximately $\sigma = \sqrt{\sigma_{meas}{}^2 + \sigma_{spatial}{}^2}$. If either standard deviation is one-third or less of the other, there is not much point in trying to reduce it further. If the smaller contributor were eliminated entirely, at most σ would be reduced by a factor of $\sqrt{9/10} \approx 0.95$, i.e., only about a 5% gain. Efforts should be directed at reducing the dominant component of the data variability.

The quality of data is critical to the successful execution of a survey. Even if the measurement uncertainty is dominated by the spatial variability, poorly calibrated instruments could lead to either improperly labeling an area as still contaminated or releasing it when, in fact, it is above the guidelines. For this reason, calibrations must be performed regularly with traceable standards; the inherent precision of the survey instrument must be evaluated to determine if it meets the needs of the survey plan. Energy responses of instruments must be known so that appropriate applications are made to different radiation fields. Replicate, reference, and blank measurements are also an integral part of the survey methodology. Comparisons of field measurement results to those of laboratory sample analyses forms an important quality control check.

Bounds on measurement uncertainties should be established in the planning process and regularly assessed throughout the measurement program. Uncertainties in the measurements add to the variance in distribution of data sets and should be taken into consideration when selecting parameters for the statistical tests and in the interpretation of results of these tests. Failure to adequately consider the effect of measurement errors could result in the added expense of additional measurements. In the worst case, inadequate control of the Type II statistical errors as determined from a retrospective power calculation, could invalidate the final survey results and require a re-survey. For this reason, it is better to plan the surveys cautiously:

- It is better to overestimate the potential data variability than to underestimate it.
- It is better to take too many samples than too few.
- It is better to overestimate minimum detectable concentrations (MDCs) than to underestimate them.

Further information on quality assurance for environmental data may be found in EPA QA/R-5 (1994), EPA QA/G-5 (1996), and ANSI/ASQC (1994)

4.2 Conduct a Preliminary Data Review

To learn about the structure of the data—identifying patterns, relationships, or potential anomalies—one can review quality assurance (QA) and quality control (QC) reports, prepare graphs of the data, and calculate basic statistical quantities.

Radiological survey data are usually obtained in units that have no intrinsic meaning relative to DCGLs, such as the number of counts per unit time. For comparison of survey data to DCGLs, the survey data from field and laboratory measurements should be converted to DCGL units.

4.2.1 Basic Statistical Quantities

Basic statistical quantities that should be calculated for the sample data set are the

- mean
- standard deviation
- median

The average of the data can be compared to the reference area average and the $DCGL_W$ to get a preliminary indication of the survey unit status. Where remediation is inadequate, this comparison may readily reveal that a survey unit contains excess residual radioactivity—even before applying statistical tests. For example, if the average of the data exceeds the $DCGL_W$ and the radionuclide of interest does not appear in background, then it is obvious that the survey unit does not meet the release criterion. On the other hand, if every measurement in the survey unit is below the $DCGL_W$, the survey unit will always pass the Sign test.

The value of the sample standard deviation is especially important. If too large compared to that assumed during the survey design, this may indicate an insufficient number of samples were collected to achieve the desired test power.

The median is the middle value of the data set when the number of data points is odd, and is the average of the two middle values when the number of data points is even. Thus 50% of the data points are above the median, and 50% are below the median. Large differences between the mean and the median would be an early indication of skewness in the data. This would also be evident in a histogram of the data.

Table 4.1 lists an example of concentration data taken in a reference area and survey unit. For this example, the quantity and units of measurement have been left arbitrary. Basic statistical quantities can be calculated simply by using one of the many widely available personal computer programs that perform data analysis. Table 4.2 shows the result of a “descriptive statistics” command applied to the data of Table 4.1 using a spreadsheet program. In addition to the mean, median and standard deviation, this table lists several other useful parameters such as the minimum, maximum, mode, range, skewness and kurtosis.

For the example survey unit, the mean is 1.15 and the median is 1.05. The sample standard deviation is 0.46. The difference between the median and the mean, divided by the sample standard deviation is sometimes used as a simple measure of skewness. In this case, (1.15 − 1.05)/0.46 = 0.22.

The coefficient of skewness is the average cubed difference from the mean divided by the standard deviation cubed. The sample estimate of skewness is $g_1 = m_3 / m_2^{3/2}$, where

$$m_3 = \sum_{i=1}^{n} (x_i - \bar{x})^3 / n \text{ , and } \quad m_2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 / n . \quad m_2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 / n .$$

For a sample from a normal distribution, g_1 is approximately normal with mean zero and standard deviation $\sqrt{6/n}$. The sample skewness for the survey unit data is 0.96. This is nearly four times $\sqrt{6/90}$ = 0.26, indicating that there is some positive skewness in this data set.

Table 4.1 Example Final Status Survey Data

Reference Area							Survey Unit					
Point No.	**Data Value**	**Point No.**	**Data Value**	**Point No.**	**Data Value**		**Point No.**	**Data Value**	**Point No.**	**Data Value**	**Point No.**	**Data Value**
1	1.1	31	1.9	61	0.8		91	1.2	121	0.7	151	0.8
2	1.3	32	1	62	1.1		92	1.4	122	1.9	152	1.1
3	0.7	33	0.7	63	0.6		93	0.8	123	1.3	153	1.2
4	0.7	34	1.9	64	0.8		94	0.6	124	1.4	154	0.7
5	1.6	35	1.0	65	1.2		95	1.4	125	0.5	155	1.4
6	1.0	36	0.6	66	0.8		96	2.9	126	1.0	156	1.6
7	1.1	37	1.5	67	1.0		97	0.9	127	1.3	157	0.4
8	0.7	38	1.1	68	0.9		98	0.8	128	0.6	158	0.6
9	0.9	39	0.9	69	1.5		99	0.8	129	1.3	159	1.6
10	1.8	40	0.9	70	0.8		100	1.6	130	1.5	160	0.7
11	0.9	41	0.8	71	1.2		101	1.6	131	1.4	161	1.0
12	0.7	42	1.1	72	1.1		102	1.2	132	1.3	162	1.0
13	1.1	43	0.9	73	0.6		103	1.2	133	0.8	163	1.8
14	1.1	44	1.2	74	1.0		104	2.5	134	1.5	164	1.3
15	0.9	45	1.2	75	0.9		105	1.9	135	0.8	165	1.5
16	1.5	46	1.0	76	1.0		106	1.9	136	1.1	166	0.8
17	1.0	47	1.3	77	0.6		107	0.9	137	1.1	167	1.5
18	0.8	48	0.9	78	0.9		108	0.9	138	1.0	168	0.9
19	0.6	49	0.8	79	1.0		109	0.8	139	1.1	169	0.9
20	1.1	50	1.7	80	0.8		110	1.0	140	1.6	170	0.8
21	0.7	51	0.7	81	0.6		111	1.7	141	1.5	171	1.5
22	0.6	52	1.0	82	1.2		112	1.5	142	0.8	172	1.0
23	0.9	53	0.8	83	1.2		113	2.1	143	0.7	173	0.7
24	0.6	54	1.0	84	1.3		114	2.0	144	0.6	174	1.1
25	1.5	55	0.5	85	1.0		115	1.7	145	0.9	175	1.4
26	0.9	56	1.1	86	0.9		116	0.7	146	0.8	176	1.0
27	1.5	57	1.1	87	0.7		117	1.0	147	0.5	177	1.2
28	0.8	58	0.9	88	0.9		118	1.0	148	0.6	178	0.5
29	1.1	59	0.9	89	1.4		119	1.5	149	0.8	179	0.5
30	1.2	60	0.6	90	1		120	1	150	0.8	180	1.7

Table 4.2 Basic Statistical Quantities Calculated for the Data in Table 4.1

Reference		Survey Unit	
Mean	1.00	Mean	1.15
Standard Error	0.03	Standard Error	0.05
Median	1.00	Median	1.05
Mode	0.90	Mode	0.80
Standard Deviation	0.30	Standard Deviation	0.46
Sample Variance	0.09	Sample Variance	0.22
Kurtosis	0.93	Kurtosis	1.44
Skewness	0.95	Skewness	0.96
Range	1.4	Range	2.5
Minimum	0.5	Minimum	0.4
Maximum	1.9	Maximum	2.9
Count	90	Count	90

The kurtosis is the average fourth power of the difference from the mean divided by the variance squared. It is a measure of how "flat" the distribution is relative to normally distributed data. The sample estimate of kurtosis is $b_2 = m_4 / m_2^2$, where

$$m_4 = \sum_{i=1}^{n} (x_i - \bar{x})^4 / n \ , \text{ and } \quad m_2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 / n = (n-1) s^2 / n.$$

For a sample from a normal distribution, b_2 has mean three. The *sample coefficient of kurtosis* is $g_2 = b_2 - 3$. For very large samples from a normal distribution, g_2 has mean zero and standard deviation $\sqrt{24/n}$ (Snedecor and Cochran, 1980). The sample coefficient of kurtosis for the survey unit data is 1.44, and .Thus, the kurtosis appears to be significantly greater than zero. It is an indicator of how well the sample variance, s^2, estimates the true variance, σ^2, of the measurement data. The variance of the sample estimate of the variance is

$$\text{Var}(s^2) = \frac{2\sigma^4}{n-1} \left[1 + \frac{n-1}{2n} g_2 \right]$$

The variance of s is given approximately by

$$\mathrm{Var}(s) \;=\; \mathrm{Var}(\sqrt{s^2}) \;\approx\; \frac{1}{4s}\;\mathrm{Var}(s^2) \;=\; \frac{2\sigma^4}{4s(n-1)}\left[1+\frac{n-1}{2n}g_2\right] \;\approx\; \frac{s^3}{2(n-1)}\left[1+\frac{n-1}{2n}g_2\right],$$

where the propagation of error formula for the variance of the square root has been used (Taylor, 1990). For the example survey unit data with s = 0.36, g_2 = 1.44, and n = 90, we have

$$\mathrm{Var}(s) \;\approx\; \frac{(0.46)^3}{2(89)}\left[1+\frac{89}{180}1.44\right] = (0.0973/178)\,(1.712) = 0.000936.$$

The standard deviation of s = $\sqrt{\mathrm{Var}(s)} \;=\; \sqrt{0.000936} \;=\; 0.0306$. Thus, one can estimate that, *very roughly*, s = 0.46 ± 0.03.

An approximate $1-\alpha$ confidence interval for σ^2 may be obtained from

$$\left[\frac{(n-1)s^2}{1+g_2/n}\right] \Big/ \; \chi^2_{n-1}(1-\alpha/2) \;<\; \sigma^2 \;<\; \left[\frac{(n-1)s^2}{1+g_2/n}\right] \Big/ \; \chi^2_{n-1}(\alpha/2)$$

where $\chi^2_{n-1}(a)$ is the 100ath percentile of the chi-squared distribution with $n-1$ degrees of freedom (Box, 1953). Percentiles of the chi-squared distribution are given in Table A.5. With s^2 = 0.22, g_2 =1.44, n = 90, $\chi^2_{89}(0.975)$ = 117, and $\chi^2_{89}(0.025)$ = 64.8, we find that with 95% confidence $\left[\frac{(89(0.22)^2}{1+1.44/90}\right] \Big/ \; 117 \;<\; \sigma^2 \;<\; \left[\frac{(89)(0.22)^2}{1+1.44/90}\right] \Big/ \; 64.8$ or $0.1647 < \sigma^2 < 0.2974$, which implies that $0.406 < \sigma < 0.545$. This is not too different from the cruder estimate $s = 0.46 \pm (2)\,(0.03) = 0.46 \pm 0.06$, using a 2σ interval about the mean.

Examining the minimum, maximum, and range of the data may provide additional useful information. The minimum of the example survey unit data is 0.4 and the maximum is 2.9, so the range is 2.9 − 0.4 = 2.5. This is 5.4 times the standard deviation of 0.46. Figure 4.1 shows that is well within the expected spread of values for this ratio, which is sometimes called the *studentized* range. These intervals were calculated for normally distributed data.

Absolute upper and lower bounds for the studentized range have been found by Thomson (1955). These bounds are fairly wide, but are useful in checking for errors in calculation. The upper bound is $\sqrt{2(n-1)}$. The lower bound is $2\sqrt{(n-1)/n}$, when n is even, and $2\sqrt{n/(n+1)}$, when n is odd.

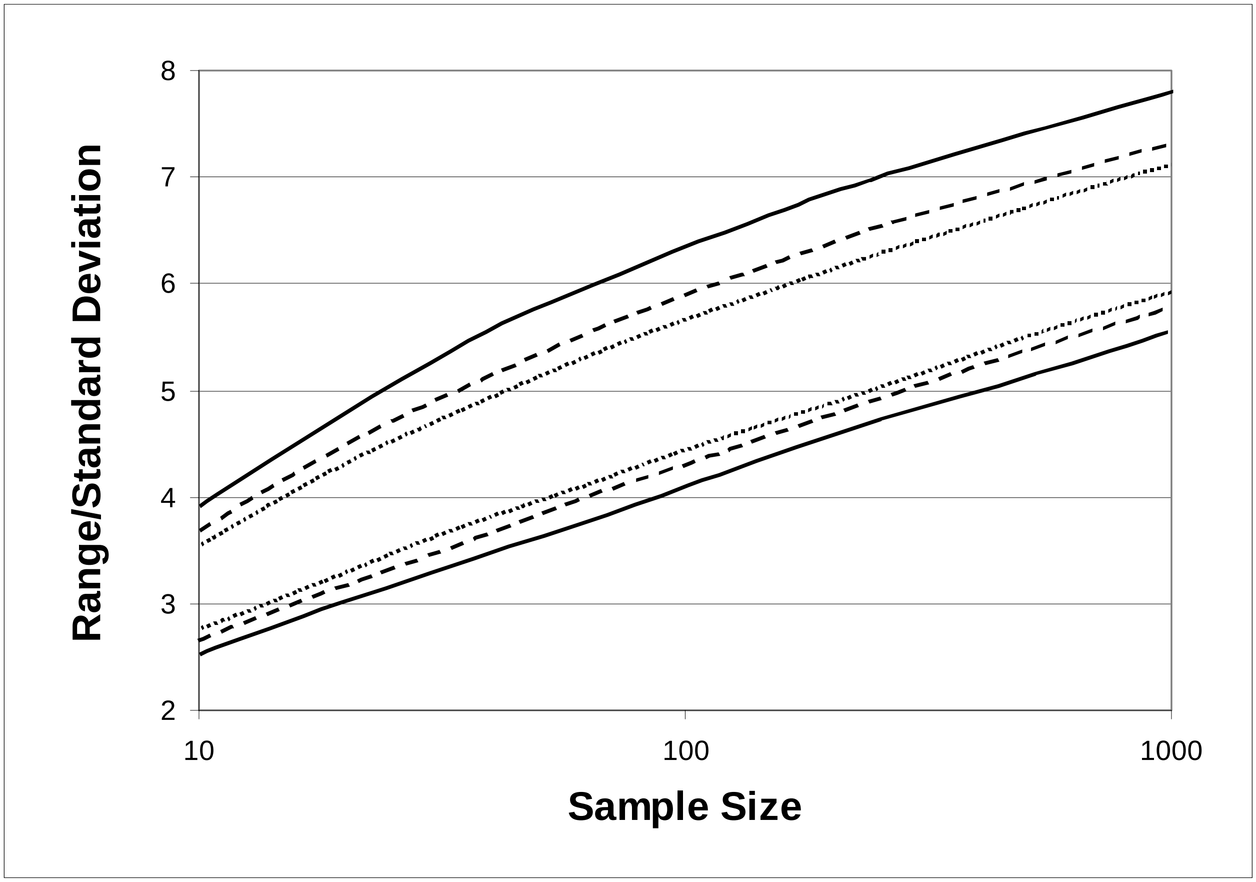

Figure 4.1 Confidence Bands for the Ratio of the Range to the Standard Deviation
Dotted: 90% Dashed: 95% Solid: 99%

Transformations are sometimes used bring data closer to a normal distribution, and decrease any dependence of the variance on the mean. A rule of thumb sometimes used is that if the ratio of the data maximum to the data minimum is less than 20, no data transformation is necessary to stabilize the variance of the data (EPA 600/4-90/013, 1990). For the example survey unit data, this ratio is 2.9/0.4 = 7.25.

Many of the "diagnostic checks" on the basic statistical quantities discussed in this section are based on comparing the values computed for the sample data distribution to those that would be expected if the data were normally distributed. When viewed as tests of normality, they are generally not very powerful, and are not suggested here for that purpose. As noted earlier, the nonparametric statistical tests used in this report do not assume the data are normally distributed. Rather, these checks are being used as exploratory techniques to alert the data analyst of any unusual features in the data.

4.2.2 Graphical Data Review

At a minimum, the graphical data review should consist of a posting plot and a histogram. Rank or Quantile plots are also useful diagnostic tools, particularly in the two-sample case, to compare the survey unit and reference area.

A *posting plot*, which is simply a map of the survey unit with the data values entered at the measurement locations, will reveal potential anomalies in the data, especially possible patches of elevated residual radioactivity. Even in a reference area, a posting plot can reveal spatial trends in background data that might affect the results of the two-sample statistical tests.

The survey unit data in Table 4.1 were taken on a square systematic grid in a rectangular survey unit. A simple posting plot is shown in the upper half of Figure 4.2. It is often useful to add some color coding of data values to aid in identifying patterns. In the lower half of Figure 4.2, darker shading was used for larger data values. The small slightly elevated region near 40 East and 20 North stands out more clearly when the shading is added.

If the posting plot reveals systematic spatial trends in the survey unit, the cause would need to be investigated. In some cases, such trends could be due to residual radioactivity, but may also be due to an inhomogeneous survey unit background. Other diagnostic tools for examining spatial data trends may be found in EPA Report QA/G-9 (1996). Geostatistical tools may also be useful in some cases (EPA 230/02-89-042, 1989a).

A *frequency plot* (or a histogram) is a useful tool for examining the general shape of a data distribution. This plot is a bar chart of the number of data points within a certain range of values. The frequency plot will reveal any obvious departures from symmetry, such as skewness or bimodality (two peaks), in the data distributions for the survey unit or reference area. Skewness or other asymmetry can impact the accuracy of the statistical tests. A data transformation (e.g., taking the logs of the data) can sometimes be used to make the distribution more symmetric. The statistical tests could then be performed on the transformed data. The interpretation of the results, however, can be more complex, since the quantity being tested is also transformed. For example, the mean of log-transformed data is the log of the geometric mean of the data, not the log of the arithmetic mean of the data.

The presence of two peaks in the survey unit frequency plot may indicate the existence of isolated areas of residual radioactivity. In some cases it may be possible to determine an appropriate background for the survey unit using this information. The interpretation of the data for this purpose will generally be highly dependent on site-specific considerations and should only be pursued after consultation with the responsible regulatory agency.

The presence of two peaks in the reference area frequency plot may indicate a mixture of background concentration distributions due to different soil types, construction materials, *etc.* The greater variability in the data due to the presence of such a mixture will reduce the power of the statistical tests to detect an adequately remediated survey unit. These situations should be avoided whenever possible by carefully matching the reference areas to the survey units, and choosing survey units with homogeneous backgrounds.

A major concern in constructing a histogram or frequency plot is the bin width, i.e. the range of concentration values over which the data are grouped and counted. If the bin width is too narrow, there will be too much spurious detail in the plot. If the bin width is too wide, too much detail is lost. A useful rule of thumb is to calculate the bin width by rounding down the quantity $3.5sn^{-1/3}$, where n is the number of data points, and s is the sample standard deviation (Scott, 1979). An example is shown in Figure 4.3 using the example survey unit data. In this example,

$3.5sn^{-1/3} = 3.5(0.46)(90)^{-1/3} = 3.5(0.32)(0.22) = 0.354$, which was rounded down to 0.3. The resulting histogram is shown as Figure 4.3a. For comparison, histograms constructed using bin widths of 0.2 (Figure 4.3b) and 0.1 (Figure 4.3c) are also shown.

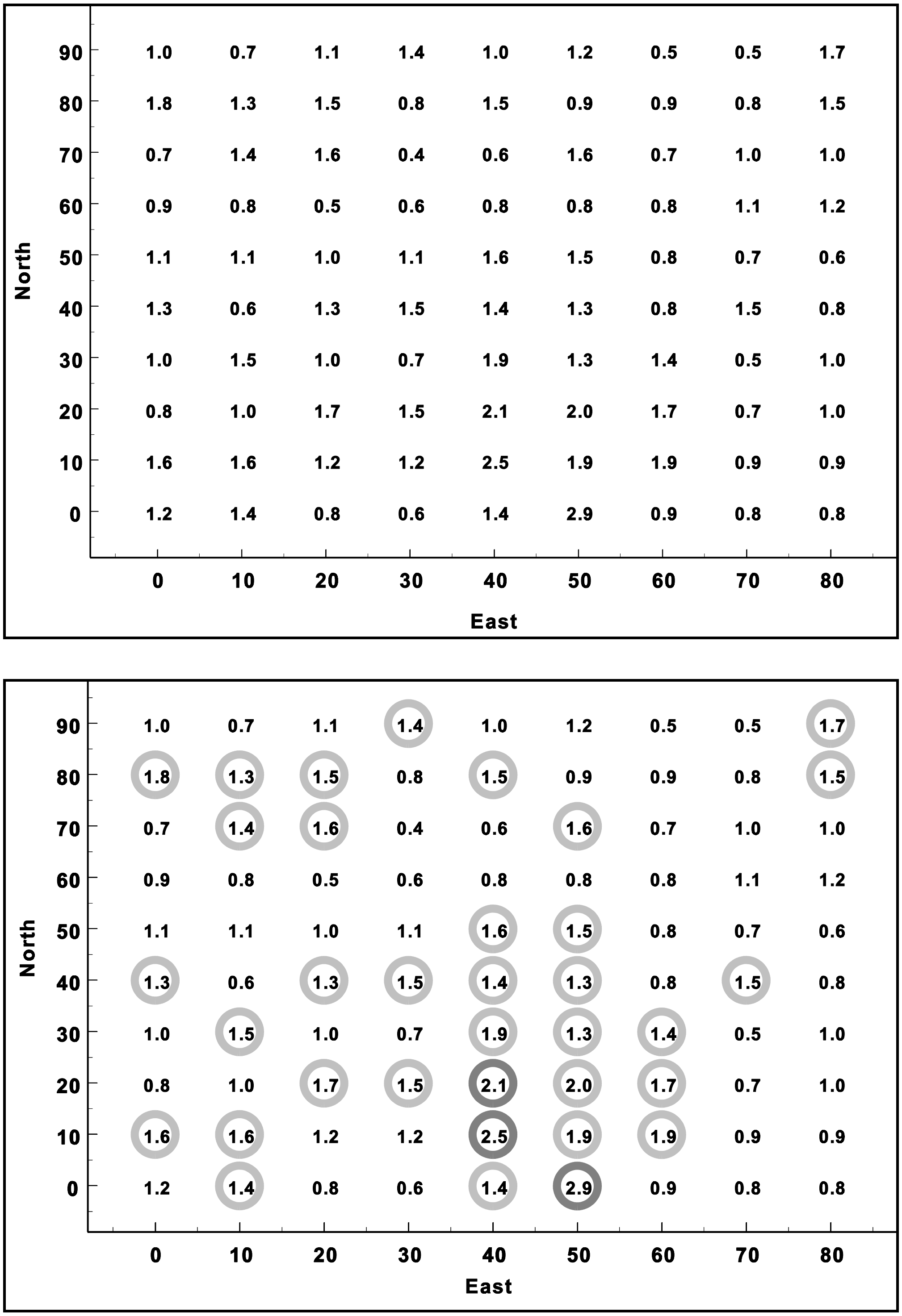

Figure 4.2 Example of a Posting Plot

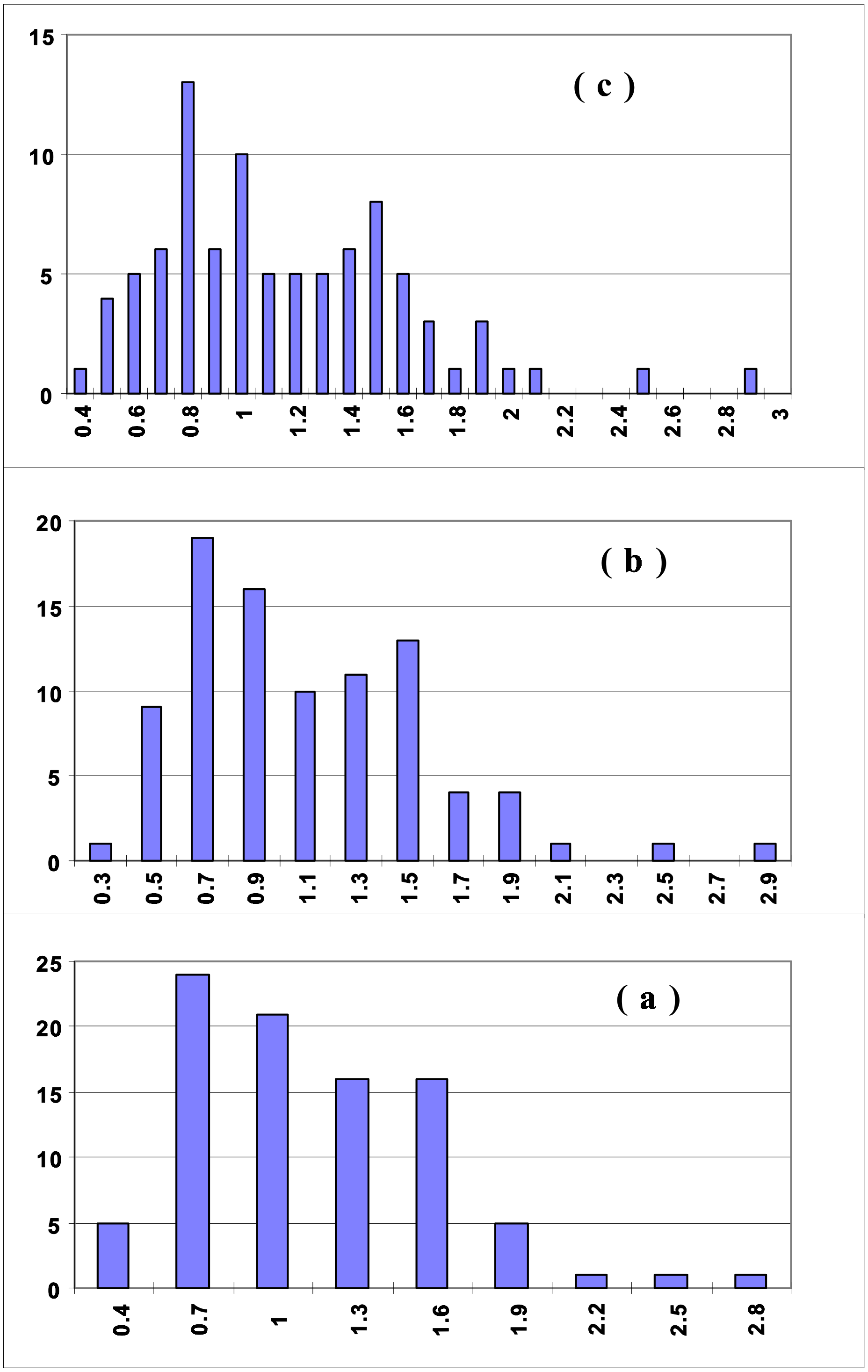

Figure 4.3 Frequency Plots of Example Survey Unit Data
with bin width 0.1 (c-top) and 0.2 (b-middle) and 0.3 (a-bottom)

A *ranked data plot* indicates the amount of data falling within a given range of values. The first step in constructing a ranked data plot is to sort the data in increasing order. The data are then assigned the number corresponding to their position in the list. The ranking of the example data from Table 4.1 is shown in Table 4.3.

Table 4.3 Ranks of the Example Data

Reference Area							Survey Unit					
Rank	**Data**	**Rank**	**Data**	**Rank**	**Data**		**Rank**	**Data**	**Rank**	**Data**	**Rank**	**Data**
1	0.5	31	0.9	61	1.1		1	0.4	31	0.9	61	1.4
2	0.6	32	0.9	62	1.1		2	0.5	32	0.9	62	1.4
3	0.6	33	0.9	63	1.1		3	0.5	33	0.9	63	1.4
4	0.6	34	0.9	64	1.1		4	0.5	34	0.9	64	1.4
5	0.6	35	0.9	65	1.1		5	0.5	35	0.9	65	1.4
6	0.6	36	0.9	66	1.1		6	0.6	36	1.0	66	1.4
7	0.6	37	0.9	67	1.1		7	0.6	37	1.0	67	1.5
8	0.6	38	0.9	68	1.1		8	0.6	38	1.0	68	1.5
9	0.6	39	0.9	69	1.1		9	0.6	39	1.0	69	1.5
10	0.6	40	0.9	70	1.2		10	0.6	40	1.0	70	1.5
11	0.7	41	0.9	71	1.2		11	0.7	41	1.0	71	1.5
12	0.7	42	0.9	72	1.2		12	0.7	42	1.0	72	1.5
13	0.7	43	0.9	73	1.2		13	0.7	43	1.0	73	1.5
14	0.7	44	0.9	74	1.2		14	0.7	44	1.0	74	1.5
15	0.7	45	1.0	75	1.2		15	0.7	45	1.0	75	1.6
16	0.7	46	1.0	76	1.2		16	0.7	46	1.1	76	1.6
17	0.7	47	1.0	77	1.3		17	0.8	47	1.1	77	1.6
18	0.7	48	1.0	78	1.3		18	0.8	48	1.1	78	1.6
19	0.8	49	1.0	79	1.3		19	0.8	49	1.1	79	1.6
20	0.8	50	1.0	80	1.4		20	0.8	50	1.1	80	1.7
21	0.8	51	1.0	81	1.5		21	0.8	51	1.2	81	1.7
22	0.8	52	1.0	82	1.5		22	0.8	52	1.2	82	1.7
23	0.8	53	1.0	83	1.5		23	0.8	53	1.2	83	1.8
24	0.8	54	1.0	84	1.5		24	0.8	54	1.2	84	1.9
25	0.8	55	1.0	85	1.5		25	0.8	55	1.2	85	1.9
26	0.8	56	1.0	86	1.6		26	0.8	56	1.3	86	1.9
27	0.8	57	1.0	87	1.7		27	0.8	57	1.3	87	2.0
28	0.8	58	1.1	88	1.8		28	0.8	58	1.3	88	2.1
29	0.9	59	1.1	89	1.9		29	0.8	59	1.3	89	2.5
30	0.9	60	1.1	90	1.9		30	0.9	60	1.3	90	2.9

The ranked data plots for this data are shown in Figure 4.4 and Figure 4.5. A small amount of data in a range will result in a large slope. A large amount of data in a range of values will result in a flatter slope. A sharp rise near the bottom or the top is an indication of asymmetry. In Figure 4.4, the is an indication of some slight asymmetry in the reference area data. There is stronger evidence of asymmetry in the survey unit data in Figure 4.5.

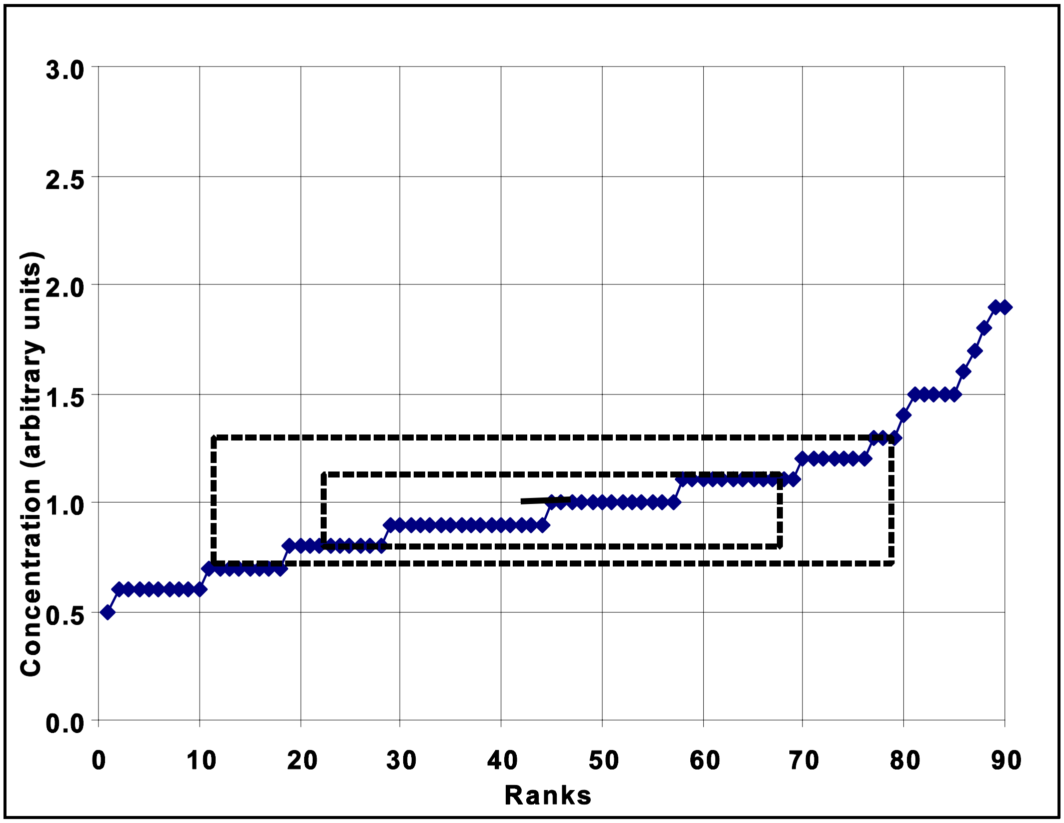

Figure 4.4 Ranked Data Plot for the Example Reference Area Data

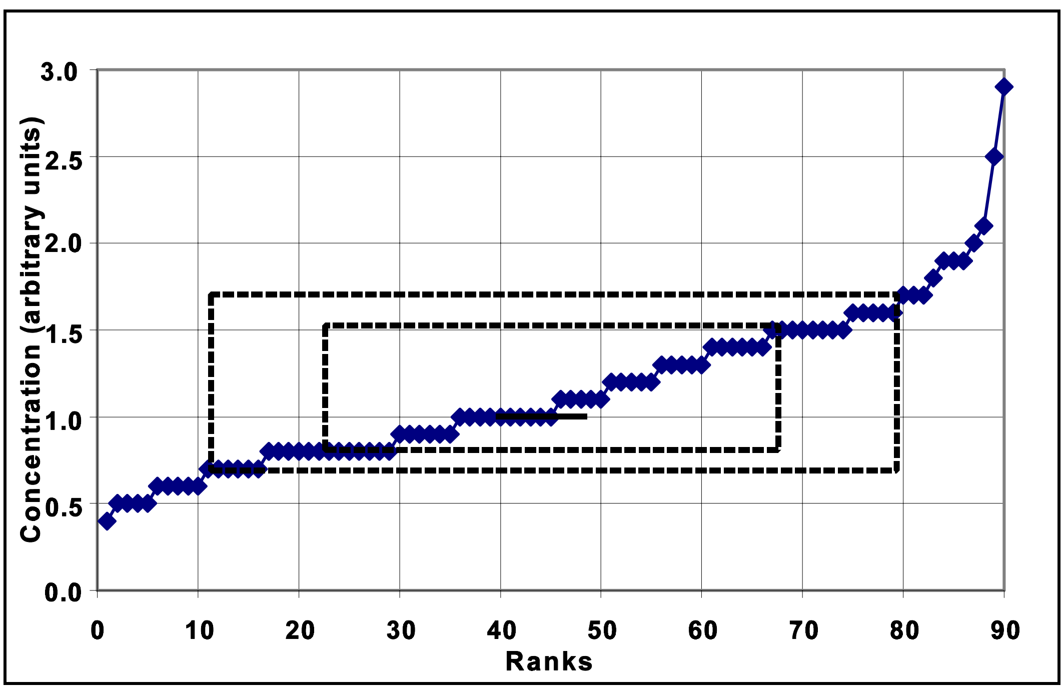

Figure 4.5 Ranked Data Plot for the Example Survey Unit Data

A Quantile plot is similar to a ranked data plot. It is constructed by ranking the data from smallest to largest, and simply plotting the data against the quantity: (rank−0.5)/(number of data points) rather than against the ranks. In this way, the percentage of data in various concentration ranges is easily found.

A useful aid to interpreting a ranked data or quantile plot is the addition of boxes containing the middle 50% and middle 75% of the data. These are shown as the dashed lines in Figure 4.4. The 50% box has its upper right corner at the 75th percentile and its lower left corner at the 25th percentile. These points are also called the quartiles. For the example survey unit data, these are 0.8 and 1.5, respectively. as indicated by the inner dashed box. They bracket the middle half of the data values. The 75% box has its upper right corner at the 87.5th percentile and its lower left corner at the 12.5th percentile. A sharp increase within the 50% box can indicate two or more modes in the data. Outside the 75% box, sharp increases can indicate outliers. The median (50th percentile) is indicated by the heavy solid line at the value 1.0, and can be used as an aid to judging the symmetry of the data distribution.

A Quantile-Quantile plot is valuable because it provides a direct visual comparison of the two data sets. If the two data distributions differ only in location (e.g., mean) or scale (e.g., standard deviation), the points will lie on a straight line. If the two data distributions being compared are identical, all of the plotted points will lie on the line $Y = X$. Any deviations from this would point to possible differences in these distributions. A Quantile-Quantile plot can be constructed to compare the distribution of the survey unit data with the distribution of the reference area data. If the number of data points is the same in both sets, the construction of the Quantile-Quantile plot is straightforward. This has already been done for the example data in Table 4.3. Simply plot each pair of measurements matched with the same rank, i.e. the survey unit measurement, *Y*, with rank *R* is plotted against the reference area measurement, *X*, with rank *R*. If the number of data points in the survey unit and reference area are not equal, the construction of the Quantile - Quantile plot will involve some numerical adjustments of the ranks. This and other useful techniques for exploratory data analysis are discussed in EPA QA/G-9 (1996).

The Quantile-Quantile plot for the example data is shown in Figure 4.6. The middle data point plots the median of the survey unit data against the median of the reference area data. That this point lies above the line $Y = X$, shows that the median of *Y* is larger than the median of *X*. Indeed, the most of the points lie above the line $Y = X$ in the region of the plot beyond a concentration value of about one. This is a sensitive indication that the distribution of the survey unit data is shifted toward values higher than the reference area distribution. As with the quantile plot, the addition of boxes containing the middle 50% and middle 75% of the data can be a useful aid to interpreting a quantile-quantile plot.

4.3 Select the Statistical Test

An overview of the statistical considerations important for final status surveys appears in Section 2.3, 3.7, and 3.8. The statistical tests recommended in this report for final status surveys are discussed in Section 2.4. The detailed instructions for applying these tests, with examples, appear in Chapters 5, 6, and 7.

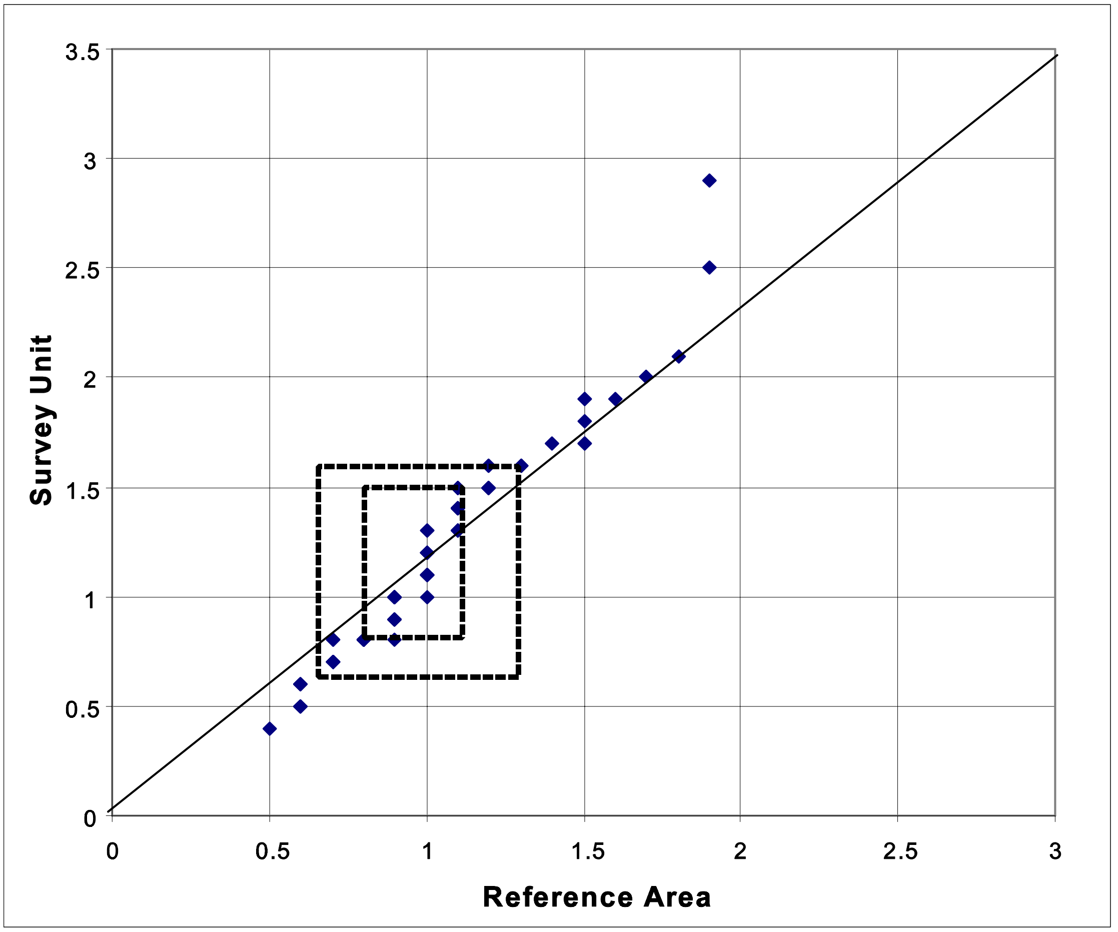

Figure 4.6 Example Quantile-Quantile Plot

The nonparametric statistical tests in this report are described as *one-sample* (Sign) and *two-sample* (WRS, Quantile) tests. Their application will depend upon the specific radionuclides under consideration, the concentration or surface activity limits for these radionuclides, and the comparison to background levels in the surrounding environment. Application of these techniques will also depend upon whether a gross dose or count rate survey is employed instead of spectrometric measurements for individual nuclides.

The one-sample tests are appropriate when there is no need to compare the survey unit with a reference area. The one-sample statistical test (Sign test) described in Chapter 5 can be used if the contaminant is not present in background and radionuclide-specific measurements are made. The one-sample test may also be used if the contaminant is present at such a small fraction of the $DCGL_W$ value as to be considered insignificant. In this case no provision for background concentrations of the radionuclide is made. Thus, the total concentration of the radionuclide is compared to the release criterion. This option should only be used if it is expected that ignoring the background concentration will not significantly affect the decision on whether or not the survey unit meets the release criterion. The advantage of ignoring a small background contribution is that no reference area is needed. This can simplify the final status survey considerably.

The two-sample WRS test (discussed in Chapter 6) should be used when the radionuclide of concern appears in background or if measurements are used that are not radionuclide specific.

The two-sample Quantile test discussed in Chapter 7 is used only when the null hypothesis of Scenario B is chosen.

The statistical tests recommended in this report are listed in Table 4.4. In every case, these tests are supplemented by the elevated measurements comparison (cf. Sections 2.6, 3.7.2, 3.8.2 and Chapter 8). Other statistical tests may be used provided that the data are consistent with the assumptions underlying their use, as discussed in the next section. The nonparametric tests generally involve fewer assumptions than their parametric equivalents. For example, the Student's *t* test may be used if the data distribution is consistent with the assumption of normality. If the data do not exhibit a normal distribution, the nonparametric tests will generally produce smaller decision error rates.

Table 4.4 Recommended Statistical Tests

Scenario	Reference Area	Test
A	Yes	Wilcoxon Rank Sum
A	No	Sign
B	Yes	Wilcoxon Rank Sum, Quantile
B	No	Sign

4.4 Verify the Assumptions of the Statistical Test

An evaluation should be made to determine that the data are consistent with the underlying assumptions of the statistical testing procedures used. Certain departures from these assumptions may be acceptable when given the actual data and other information about the study. Much of the information gained in the preliminary data review (Section 4.2) is directly applicable to verifying the assumptions of the statistical tests, and is a major reason for emphasizing their use.

A statistical test is called robust if it relatively insensitive to departures from its underlying assumptions. The nonparametric procedures described in this report were chosen because they are robust for the problem of testing the value of mean concentrations of residual radioactivity in a survey unit. In cases where the data distributions are extremely skewed, these tests may not detect limited areas with concentration much higher than the average in the survey unit. This is one reason for supplementing these tests with the elevated measurement comparison.

The nonparametric tests of Chapters 5, 6 and 7 assume that the data from the reference area or survey unit consist of independent samples from each distribution. The WRS test assumes that the reference area and survey unit data distributions are the same except for a possible shift in the mean.

Spatial dependencies that potentially affect the assumptions can be assessed using the posting plots. More sophisticated tools for determining the extent of spatial dependencies are also available (e.g., EPA QA/G-9, 1996). These methods tend to be complex and are best used with guidance from a professional statistician.

Asymmetry in the data can be diagnosed with a histogram or a ranked data plot. Data transformations can sometimes be used to minimize the effects of asymmetry.

One of the primary advantages of the nonparametric tests used in this report is that they involve fewer assumptions about the data than their parametric counterparts. If parametric tests are used, (e.g., Student's *t*-test), then any additional assumptions made in using them should be verified (e.g., testing for normality). These issues are discussed in detail in EPA QA/G-9 (1996).

Some alternative tests that may be considered in certain situations are discussed in Chapter 14. For example, if the data are symmetric, the one-sample WSR test is generally more powerful than the Sign test.

Table 4.5 Methods for Checking the Assumptions of Statistical Tests

Assumption	Diagnostic
Spatial Independence	Posting Plot
Symmetry	Histogram, Quantile Plot, Skewness
Data Variance	Sample Standard Deviation, Kurtosis
Power is Adequate	Retrospective Power Chart

4.5 Draw Conclusions From the Data

Perform the calculations required for the statistical tests and document the inferences drawn as a result of these calculations. The specific details for conducting the statistical tests are given in Chapters 5, 6 and 7.

In each survey unit, there are two types of measurements made: (1) direct measurements or samples at discrete locations and (2) scans. The statistical tests are only applied to the measurements made at discrete locations. When the data clearly show that a survey unit meets or exceeds the release criterion, the result is often obvious without performing the formal statistical analysis. Table 2.3 in Section 2.5 discussed those circumstances where a conclusion can be drawn from a simple examination of the data.

Sections 2.5.6 and 2.5.7 discuss the elevated measurement comparison (EMC) and the investigation levels that flag a locations for further study in order to determine whether the survey unit meets or exceeds the release criterion.

This report has been fairly explicit about the steps that should be taken to show that a survey unit meets release criteria. Less has been said about the procedures that should be used if at any point the survey unit fails. This is primarily because there are many different ways that a survey unit may fail the final status survey. The overall level of residual radioactivity may not pass the nonparametric statistical tests. Further investigation following the elevated measurement comparison may show that there is a large enough area with a concentration too high to meet the dose criterion. Investigation levels may have been flagged during scanning that indicate unexpected levels of residual radioactivity for the survey unit classification. It is impossible to enumerate all of the possible reasons for failure, their causes, and their remedies.

When a survey unit fails the release criterion, the first step is to review and confirm the data that led to the decision. Once this is done, the extent of the residual radioactivity is that caused the failure should be determined. Once the cause of failure has been remediated, determine the additional data, if any, needed to document that the survey unit meets the release criterion.

For example, a Class 2 survey unit passes the nonparametric statistical tests, but has several measurements on the sampling grid that exceed the $DCGL_W$. This is unexpected in a Class 2 area, and according to Table 2.4, these measurements are flagged for further investigation. Additional sampling confirms that there are several areas where the concentration exceeds the $DCGL_W$. This indicates that the survey unit was mis-classified. However, the scanning technique that was used was sufficient to detect residual radioactivity at the $DCGL_{EMC}$ calculated for the sample grid. No areas exceeding the $DCGL_{EMC}$ where found. Thus, the only difference between the final status survey actually done, and that which would be required for Class 1, is that the scanning may not have covered 100% of the survey unit area. In this case, it would be reasonable to simply increase the scan coverage to 100%. If no areas exceeding the $DCGL_{EMC}$ are found, the survey unit has, in effect, met the release criteria as a Class 1 survey unit.

A second example might be a Class 1 Survey unit which passes the nonparametric statistical tests, but which contains some areas that were flagged for investigation during scanning. Further investigation, sampling and analysis indicates one area is truly elevated. This area has a concentration that exceeds the $DCGL_W$ by a factor greater than the area factor calculated for its actual size. This area is remediated, and remediation control sampling shows that the residual radioactivity was removed, and no other areas were contaminated with removed material. It may be reasonable in that case, to simply document the original survey and the additional remediation data. It is not clear that further final status survey data would provide any useful information.

As a last example, consider a Class 1 area which fails the nonparametric statistical tests. Confirmatory data indicates that the average concentration in the survey unit does exceed the $DCGL_W$ substantially over a majority of its area. There would appear to be little alternative to remediation of the entire survey unit, followed by another final status survey.

These examples are meant to illustrative of the actions that may be necessary to secure the release of a survey unit that has initially failed to meet the release criterion. The DQO process should be revisited so that a plan can be made for attaining the original objective: to safely release the survey unit by showing that it meets the release criteria. Whatever data is necessary to meet this objective will be in addition to the final status survey data already in hand. All of the data, and only the data, necessary to meet the objective should be required.